Using full genome sequence to compare pathogenic and non-pathogenic isolates of *Theileria orientalis* in Australia

A thesis submitted in fulfilment of the requirements for the award of the degree

Doctor of Philosophy

from

University of Technology Sydney

by

Jerald Guojun Yam, BSc (Hons)

School of life sciences

August 2020

Declaration

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Jerald G Yam, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 11th August 2020

Acknowledgements

I wish to express my sincere appreciation to my supervisors, Dr. Cheryl Jenkins, Dr. Daniel Bogema and Prof Steven Djordjevic for their support, patience, wisdom, guidance and the opportunity for completing this Ph.D. project. To Cheryl, thank you for the encouragement; taking the time to review my work and progress; and always providing valuable feedback. To Daniel, thank you for the support throughout my candidature, especially with bioinformatic tasks, this could not have been achieved without your guidance. To Steve, thanks for being my principal supervisor and handling my scholarships and travel grants.

I would like to thank AUSGEM for the scholarship stipend as well as the University of Technology Sydney (UTS) for the international research scholarship.

I am also grateful to Dr. Michael Liu, formally from UTS and the Ramaciotti Centre for Genomics for their Illumina sequencing services.

All of my research was conducted at the Elizabeth Macarthur Agricultural Institute (EMAI), and I would like to acknowledge Dr. Melinda Niedermayer and Shayne Fell for their technical assistance. Also, everyone in the Microbiology and Parasitology section as well as other sections at EMAI for the memorable and joyful experience.

A Ph.D. can be mentally challenging, and I would like to thank my fellow teammates from CYL dragonboat club for their encouragement, support and providing me the much-needed work-life balance.

Words cannot express how grateful I am to my parents, Yvonne and Jonathan Yam. Thank you for the sacrifices and opportunities for me to pursue my dreams. Also, to my sister, Jillian Yam, thank you for the moral support. Last but definitely not least, to my wife, Paulina Sudjatmiko-Yam, thank you for the encouragement, understanding and patience. It was definitely not an easy journey, but your unwavering support has given me the confidence to overcome many obstacles throughout my candidature.

Format of thesis

The Ph.D. thesis presented here for examination is in the format of compilation. This study includes a published introduction chapter and four research chapters in paper publication format. The rationale for this thesis format is to enable the reader to have a better flow of information. Each research chapter contributes and builds on the following chapter to achieve the various aims of this study and it also facilitates the transition to publication.

List of publications

1) Yam, J., Bogema, D.R. and Jenkins, C., 2018. Oriental Theileriosis. In: Ticks and Tick-Borne Pathogens. IntechOpen. DOI: 10.5772/intechopen.81198. *

2) Yam, J., Gestier, S., Bryant, B., Campbell-Ward, M., Bogema, D., Jenkins, C., 2018. The identification of *Theileria bicornis* in captive rhinoceros in Australia. Int. J. Parasitol. Parasites Wildl. 7, 85-89.

* Article reproduced with permission from publisher.

Table of contents

List of Figures

Supplementary data

List of Tables

Abbreviations

AD	Alternate depth
ANI	Average nucleotide identity
BAM	Binary alignment map
BWA	Burrows-Wheeler Aligner
COGS	Cluster of ortholog groups
COX	Cytochrome oxidase
CTL	Cytotoxic T lymphocytes
EDTA	Ethylenediamine tetraacetic acid
EMAI	Elizabeth Macarthur Agricultural Institute
FN	False negative
FP	False positive
GATK	Genome analysis toolkit
GCN	Gene copy number
gDNA	genomic DNA
GPU	Graphics processing unit
HPCC	High performance computer cluster
HS	High sensitivity
HiS	HiSeq
IBD	Identity by descent
IFAT	Indirect fluorescent antibody test
INDEL	Insertion and deletion
ITS	Internal transcribed spacer
LF	Leukocyte filtration
MAF	Minor allele frequency
MCMC	Markov chain Monte Carlo
MDS	Multidimensional scaling
MGW	Molecular grade water
MOPS	3-Morpholinopropane-1-sulfonic acid
MPSP	Major piroplasm surface protein
MiS	MiSeq

NCBI	National Centre for Biotechnology Information
NGM	NextGenMap
NGS	Next generation sequencing
NSW	New South Wales
ONT	Oxford Nanopore Technologies
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PLAF	Population level allele frequency
PPV	Positive predictive value
PVP	Parasite variant pipeline
QLD	Queensland
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell
RD	Reference depth
RDP	Recombination detection program
RFLP	Restriction fragment length polymorphism
RLB	Reverse line blot
RT	Room temperature
SAM	Sequence alignment map
SNP	Single nucleotide polymorphism
SWG	Selective whole genome
SWGA	Selective whole genome amplification
TE	Tris-EDTA
TN	True negative
TP	True positive
UNSW	University of New South Wales
UTS	University of Technology Sydney
VCF	Variant calling format
WA	Western Australia
WSAF	Within sample allele frequency

Abstract

Theileria orientalis is an apicomplexan haemoparasite that causes oriental theileriosis in cattle. The parasite is an economic burden globally and there are currently no suitable therapies for the disease in Australia. It is vital to understand the genome of *T. orientalis* to aid the development of efficacious vaccines and/or therapeutics. Next generation sequencing provides an opportunity to comparatively study the *T. orientalis* genome to elucidate its diversity and population structure.

Removal of contaminating host DNA is critical for next-generation sequencing (NGS) of intracellular parasites. This thesis investigated the efficacy of *T. orientalis* DNA enrichment on fresh and frozen clinical blood samples using cellulose leukocyte depletion columns, a commercial microbiome enrichment kit, and selective whole genome amplification (SWGA). All methods were demonstrated to be effective but with unique limitations. qPCR assays were developed to estimate host-to-parasite ratio and identified the commercial microbiome kit to be the most feasible enrichment method for this project.

Long read sequencing of three common *T. orientalis* genotypes was achieved with an Oxford Nanopore MinION flow-cell. Using four distinct assembly methods and manual scaffolding, chromosomal-level assemblies of all three genomes was achieved. Genome annotation and comparative analysis revealed species-level, structural and gene content differences between pathogenic (genotype Ikeda, str. Robertson) and non-pathogenic (genotypes Chitose and Buffeli, str. Fish Creek and str. Goon Nure, respectively) strains similar to that of transforming Theilerias, *T. parva* and *T. annulata*.

To generate high-quality SNPs, a Parasite Variant Pipeline (PVP) was developed with stringent quality filters to eliminate low coverage, erroneous or non-coding SNPs. Validation was performed with 270 simulated sequences over three levels of coverage

breadths and real-world data. To account for errors caused by large indel events, PVP was revalidated and validation metrics revealed no false positives as well as > 88% PVP sensitivity over 16 × coverage.

Finally, a study of diversity and recombination within clinical Australian isolates was conducted with methods and tools investigated in the preceding research chapters. Illumina sequencing was performed on clinical samples that were enriched for parasite DNA. Sequencing data was processed through PVP with long read sequenced and previously published reference sequences to detect high-quality SNPs that were used to identify signatures of meiotic recombination and genetic diversity with genomic distance measures. A total of 15,901 high-quality SNPs were detected by PVP and results indicate *T. orientalis* has low sequence diversity in Australia, which suggests potential for the development of a successful vaccine.

Statement on ethics

During the course of my candidature, there were no procedures performed on any animals. All blood samples used in the studies were collected as part of routine clinical investigations by registered veterinarians. Purified piroplasms of the Robertson, Fish Creek and Goon Nure strains were collected in a previous study (Bogema et al., 2018). That study was carried out in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes at the Tick Fever Centre, Wacol, Queensland.